Introduction to PPCA and EM Algorithm

Group 10
Akshat Doshi, Faheem Samol, Neha Aytoda, Rajvi Prajapati, Ratnesh Shah
School Of Engineering & Applied Sciences, Ahmedabad University.

***Abstract*—Principal component analysis (PCA) is a very popular method that it is widely used for dimensionality reduction in data analysis, the model parameters can be computed directly from the data. Despite these attractive features however, PCA models have several limitations. One is that naive methods for finding the principal component directions have trouble with high dimensional data or large numbers of data points. Computing the sample covariance itself is very costly and PCA doesn't deal properly with missing data. So, we demonstrate how the principal axes of a set of observed data vectors may be determined through maximum-likelihood estimation of parameters in a latent variable model closely related to factor analysis for which computationally efficient EM algorithms can be used to estimate principal subspace iteratively.**

***Keywords*—PCA, Probabilistic PCA, Probability Model, Expectation Maximization, Data Compression**

I. Introduction

Principal component analysis (PCA) is a technique used to emphasize variation and bring out strong patterns in a dataset. It is often used to make data easy to explore and visualize and it's applications in computer vision include data compression, image processing, visualization and exploratory data analysis etc. The conventional PCA has certain limitations which include that it does not have a probabilistic model for observed data and also that it does not deal with missing data. Such limitations can be addressed by deriving PCA from a probabilistic framework resulting in probabilistic PCA (PPCA). The benefit of doing so is that maximum likelihood estimates for the parameters associated with the covariance matrix can be efficiently computed from the data principal components for that we use EM algorithm.

II. Model Description

A. Factor Analysis

A latent variable model seeks to relate a d-dimensional observation vector **t** to a corresponding *q*-dimensional vector of unobserved variables **x**. The most common such model is *factor analysis* where the linear relationship can be given as:

$$t = Wx + \mu + \epsilon$$

Let μ permit that our model has a non-zero mean. ϵ is a Gaussian noise modeled as $\epsilon \sim \mathcal{N}(0, \psi)$. The latent variable x is taken as $x \sim \mathcal{N}(0, I)$. The above equation induces a corresponding Gaussian distribution for the observations $t \sim \mathcal{N}(\mu, WW^T + \psi)$.

B. Probabilistic Principal Component Analysis

It can be said to be a maximum likelihood solution of a probabilistic latent variable model. The use of isotropic Gaussian noise model $\mathcal{N}(0, \sigma^2\mathbf{I})$ for ϵ in conjunction with above equation implies that the **x** conditional probability distribution over **t**-space is given by

$$\mathbf{t}|\mathbf{x} \sim \mathcal{N}(Wx + \mu, \sigma^2\mathbf{I})$$

The marginal distribution for the observed data **t** is readily obtained by integrating out the latent variables and is likewise Gaussian: $\mathbf{t} \sim \mathcal{N}(\mu, C)$ where the observation covariance model is specified by $C = WW^T + \sigma^2 I$. Additionally, EM algorithm can be employed to estimate the parameters W and σ^2 of the PPCA model.

C. Expectation Maximization Algorithm

It is an iterative method to find maximum likelihood estimation of parameters, where the model depends on missing latent variables. Iterative process to estimate parameters consists of following two steps for each iteration until the resulting values converge to fixed points:

1) Expectation (E-step): Complete all hidden and missing variables from current set of parameters by creating a function for the expectation of the log-likelihood.

2) Maximization likelihood (M-step): Update set of parameters using MLE on expected log-likelihood found, from complete set of data from previous step.

D. EM Algorithm for PPCA

Expectation step: compute for $n = 1, ..., N$,

$$\langle X_n \rangle = M^{-1}W^T(t_n - \mu),$$
$$\langle X_n X_n^T \rangle = \sigma^2 M^{-1} + \langle X_n \rangle \langle X_n \rangle^T.$$

Maximization step: $\langle L_c \rangle$ is maximized with respect to W and σ^2 giving new parameter estimates,

$$\tilde{W} = \left[\sum_n (t_n - \mu)\langle X_n \rangle^T\right]\left[\sum_n \langle X_n X_n^T \rangle\right]^{-1}$$

$$\tilde{\sigma}^2 = \frac{1}{Nd}\sum_{n=1}^{N} ||t_n - \mu||^2 - 2\langle X_n \rangle^T \tilde{W}^T (t_n - \mu) + trace(\langle X_n X_n^T \rangle \tilde{W}^T \tilde{W}).$$

References

[1] Probabilistic Principal Component Analysis Michael E. Tipping Christopher M. Bishop, Microsoft research 1999.